

# VU Research Portal

## Stated Choice Experiments with Repeated Observations

Ouwersloot, H.; Rietveld, P.

### ***published in***

Journal of Transport Economics and Policy  
1996

### ***document version***

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### ***citation for published version (APA)***

Ouwersloot, H., & Rietveld, P. (1996). Stated Choice Experiments with Repeated Observations. *Journal of Transport Economics and Policy*, 30(6), 203-212.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# ***Stated Choice Experiments with Repeated Observations***

**By Hans Ouwersloot and Piet Rietveld\***

## **1. Introduction**

Discrete choice models are an important tool in the analysis of travel behaviour (Ortúzar and Willumsen, 1990). The estimation theory of these models is mainly based on large sample results. This means that the number of observations,  $N$ , on which estimations are based, should tend to infinity for the results to have desirable properties (such as, consistency).

Empirical researchers sometimes worry about too small a value of  $N$  and try to increase it. In the context of stated choice modelling (an approach often used in travel demand modelling; see Bates, 1988) one may proceed in two directions: increase the number of respondents, or increase the number of observations per respondent. The latter alternative is usually much cheaper than the former. However, it gives rise to specific interdependency problems in the error structure (auto-correlation) which are ignored in the standard logit or probit models usually applied in this context (see Bates, 1988). The problem is similar to the auto-correlation problem encountered in the panel data literature. This paper shows how the solution from the panel data field (see, for example, Hsiao, 1986; Chamberlain, 1984) can be applied in the context of stated choice experiments.

The method itself is not new. As far as we know, however, it has seldom (or never) been used in discrete choice models. It should, nevertheless, be realised that auto-correlation can be a serious problem. The purpose of this paper is therefore twofold: first, to demonstrate the application of the correct estimation method in the case of repeated observations on discrete choice; and second, to investigate the seriousness of the auto-correlation problem.

The model will be applied in the realm of trip generation/distribution modelling. Respondents are offered alternative hypothetical destinations and they can indicate which of these destinations are worth a visit. More specifically, the application concerns a scientist who decides what types of colleagues at various locations are worth a visit to carry out collaborative research.

The paper proceeds as follows. Section 2 states the problem in formal terms and introduces the estimation method. In Section 3 the data of our application are presented. In Section 4 the results of the estimations are given, and in Section 5 some conclusions are drawn.

---

\* Free University, Amsterdam.

## 2. The Estimations of Discrete Choice Models with Repeated Observations

Consider a stated preference (SP) experiment in which respondents have to make a discrete choice. The choice set consists of two or more hypothetical alternatives and the respondent is requested to indicate the alternative he or she prefers. In order to increase the number of observations the respondents are provided with several subsequent choice situations or cases, with varying combinations of features.

This general description of an SP experiment with repeated observations fits, for example, the experiment as carried out by Wardman (1988) who offers his respondents a choice between coach and train, varying the difference in the values for each mode of such attributes as "In-Vehicle-Time", "Walk-Time", "Wait-Time" or "Cost". A dataset analysed in full detail in Ouwersloot and Rietveld (1993) also follows the general outline of the preceding paragraph. This dataset consists of observations on respondents who were offered a contact decision. The respondents, who were scientists at universities, had to respond to an invitation of a colleague at another university to collaborate in the writing of a book. The choice set thus contained two alternatives: (a) accept the invitation; (b) reject the invitation. The case was described so that, in order to achieve the collaboration, the respondent would travel to the inviter.

The features that were varied across the different cases were the location, the reputation, the professional status and English skills of the inviter. Each respondent was requested to treat four different cases. In the next section we will discuss this contact decision problem in more detail; in this section we will concentrate on estimation issues.

The data that result from such a procedure show characteristics similar to panel data: repetition of observations generated by one person. This repetition gives rise to dependencies in the presumed error structure. The potential problems that result are well known and extensively addressed in the panel data literature.

From this field, estimation procedures have been proposed to address the problem. The following exposition can be compared with Hsiao (1986) or Chamberlain (1984).

The problem is stated formally by defining

$y_{it} = 1$  if respondent  $i$  accepts the invitation in case  $t$ ; 0 otherwise;

$x_{it}$  is the vector of exogenous variables;

$\beta$  is the vector of parameters;

$\varepsilon_{it}$  is the error term.

Then the standard discrete choice formulation is

$$y_{it} = 1 \text{ if } \beta' x_{it} + \varepsilon_{it} > 0.$$

In the present context of stated choice experimentation the term 'case' is more appropriate than the term 'period', which is used in the panel data literature, although there is conceptually no difference.

As indicated above, the standard assumption that all error terms are independent is not realistic with the present kind of data. Instead, it can be assumed that the errors are independent across respondents but not for respondents across cases. Formally:

$$\begin{aligned}
E(\varepsilon_{it}) &= 0 \text{ for all } i, t; \\
E(\varepsilon_{it}, \varepsilon_{js}) &= 0 \text{ if } i \neq j, \text{ for all } t, s; \\
&= \sigma^2 \text{ if } i = j, t = s; \\
&= \rho_{ts}\sigma^2 \text{ if } i = j, t \neq s.
\end{aligned}$$

Neglecting this correlation structure may cause serious estimation errors, as Hsiao (1986, pp.6-7) demonstrates.

This error structure is clearly too general to be of any practical use. A good solution is to decompose the error  $\varepsilon$  into two parts which are mutually independent:  $\eta_i$ , an individual specific effect, independent across  $i$ ; and  $u_{it}$ , a true error term, independent across both  $i$  and  $t$ . This results in the following formulation of the model:

$$y_{it} = 1 \text{ if } \beta' x_{it} + \eta_i + u_{it} > 0.$$

This decomposition is justified by observing that the error term  $\varepsilon_{it}$  may be assumed to capture two unobserved effects, namely, some unobserved variables related to the individual, and measurement errors, which are not specific to the individual but occur for each individual at each observation.

It should be noted that with true panel data, where the auto-correlation is driven by the time structure of the data, the correlation may crucially depend on the difference between  $t$  and  $s$ . For example, the correlation may be assumed to be larger, the closer  $t$  and  $s$  are. In the repeated choice case as discussed here, the difference between  $t$  and  $s$  is less relevant. In contrast, what may be relevant is similarity between choice situations. We return to this point in Section 4.

There are two ways to deal with  $\eta_i$ . The first is to treat this term as fixed, in which case the variable is acting as an individual specific constant  $c_i$ . The second option is to treat  $\eta_i$  as a random effect. In that case the  $\eta_i$  are assumed to be drawings from some distribution with zero expectation and variance  $\sigma_\eta^2$ .

In the fixed effects case, the individual specific constants  $c_i$  cannot be treated as just a number of additional variables unless  $T$  (the number of cases or periods in the sample) tends to infinity, which is normally not the case. Standard estimation of these coefficients  $c_i$  is likely to lead to inconsistency of the estimates  $\hat{c}_i$ , and, moreover, to inconsistent estimates of  $\beta$ ,  $\hat{\beta}$ . Therefore a complex method has to be used to solve this inconsistency problem. This method, which is, for example, described in Hsiao (1986), is known to work for logit models, but not for probit models.

In the random effects model, on the other hand, a similarly complex estimation procedure as described below is applicable to probit and not to logit.

An important observation is that in the fixed effects case the estimation method has to deal with heterogeneity which is captured in  $N$  parameters (individual specific constants), while for the random effects model there is only one parameter (the variance of the distribution of  $\eta$ ). Therefore we intuitively conjecture that the information in the sample is more efficiently used in the random effects case, where the data have to take care of only one parameter, than in the fixed effects case, and hence, unless there are clear reasons to prefer a logit model, the random effects model is advocated. Consequently, in this paper we only deal with the random effects case.

With  $\eta$  random, there are two random terms. Assume that the cumulative distribution function of  $u$  is  $F$ , and the cumulative distribution function of  $\eta$  is  $H$ , and accordingly  $f$  and  $h$  are the respective probability density functions. The presence of two random terms complicates the estimation by maximum likelihood considerably. To show this we derive the log-likelihood function.

As given above, the probability that  $y_{it} = 1$  is

$$\text{Prob}(y_{it} = 1) = \text{Prob}(\beta'x_{it} + \eta_i + u_{it} > 0).$$

This can be reformulated, on condition that  $\eta_i = \eta_i^*$ , as:

$$\text{Prob}(u_{it} > -\beta'x_{it} - \eta_i) = \text{Prob}(u_{it} > -\beta'x_{it} - \eta_i | \eta_i = \eta_i^*) * \text{Prob}(\eta_i = \eta_i^*).$$

Thus, conditional on a value for  $\eta_i$  the contribution to the likelihood function of the  $T$  observations on subject  $i$  are

$$\prod_{t=1}^T [1 - F(-\beta'x_{it} - \eta_i^*)]^{\eta_i^*} [F(-\beta'x_{it} - \eta_i^*)]^{1-\eta_i^*} h(\eta_i^*),$$

where  $h$  is the probability density function of  $\eta_i$ . Next, the conditioning is integrated out over all values of  $\eta_i$  to arrive at the contribution to the likelihood function of subject  $i$

$$\int_{\eta_i=-\infty}^{\infty} \prod_{t=1}^T [1 - F(-\beta'x_{it} - \eta_i)]^{\eta_i} [F(-\beta'x_{it} - \eta_i)]^{1-\eta_i} h(\eta_i) d\eta_i.$$

Finally, multiplying the contributions of all subjects and taking the log leads to the following log-likelihood function

$$\text{Log } L = \sum_{i=1}^N \log \int_{\eta_i=-\infty}^{\infty} \prod_{t=1}^T [1 - F(-\beta'x_{it} - \eta_i)]^{\eta_i} [F(-\beta'x_{it} - \eta_i)]^{1-\eta_i} h(\eta_i | \delta) d\eta_i$$

where the probability density function of  $\eta_i$ ,  $h$  is characterised by a vector  $\delta$  of parameters (obviously independent of  $i$ ),  $N$  is the number of respondents and  $T$  is the number of cases. Maximising this log-likelihood would, under weak regularity conditions, lead to consistent estimators of  $\beta$  and  $\delta$  as  $N$  tends to infinity. The maximisation of this log-likelihood function however, would imply the  $N$ -fold evaluation of a  $T$ -multiple integral, which is very unattractive from a computational viewpoint, although Butler and Moffitt (1982) have shown that relatively good approximations can be obtained in reasonable computer time (see also Maddala, 1987).

Chamberlain (1984) proposes an alternative method of estimating  $\beta$  which avoids numerical integration and which is based on the efficiency of binomial probit estimation. This method treats each of the  $T$  cases as separate samples, each leading to an estimate of  $\beta$ , say  $\beta_t$ , and then finds an estimate of  $\beta$  using  $\beta_t$  in a minimum distance estimation procedure.<sup>1</sup> Since each observation appears only once in each subsample, there is no longer auto-correlation in the errors when estimating each of these subsections. By treating the  $T$  subsamples independently, the autocorrelation problem has thus been transferred to the second stage of the estimation. In this second stage, the  $T$  estimates  $\beta_t$  are used to arrive at the overall estimator  $\beta$ , and in this stage the possible auto-correlation is taken into account.

<sup>1</sup> Sometimes also called minimum chi-square method, or method of moments estimation.

A prerequisite for this method to work is that for each subset, that is, for fixed  $t$ , the distribution of  $\eta_i + \varepsilon_{it}$  is known, and preferably easy to handle. A convenient choice, therefore, is to assume that both terms are normally distributed, which implies that their sum also has a normal distribution. Since  $t$  is fixed, and following from previous assumptions, the term  $\eta_i + \varepsilon_{it}$  is independent across  $i$ , and the problem reduces at this first stage to standard probit modelling.

Moreover, since the parameters are estimable up to a scale factor which depends solely on the variances of  $\eta$  and  $\varepsilon$ , it is convenient to assume that the variance of the sum  $\eta_i + \varepsilon_{it}$  equals one. Because this technique is not commonly known it is described in some detail here.

The procedure starts with the independent probit estimation of the  $T$  subsets leading to  $T$  estimators  $\beta_t$  which are vectors of length  $M$ . The  $\beta_t$ 's are used to construct a minimum distance estimator  $\beta^*$ , choosing  $\beta$  to minimise

$$[\pi^* - f(\beta)]' \Omega^{-1} [\pi^* - f(\beta)] \quad (1)$$

where

$$\begin{aligned} \pi^* &= (\beta_1' \dots \beta_T')' \\ f(\beta) &= (I_M \dots I_M)' \beta \end{aligned}$$

and where in the last definition there are  $T$  identity matrices  $I_M$ .

Any choice of a positive definite  $\Omega$  leads to some minimum distance estimator of  $\beta$ , therefore the specific choice is guided by the desired characteristics of the resulting estimator. Hsiao (1986) shows that  $\Omega$  as defined below leads to a consistent and asymptotically normal minimum distance estimator of  $\beta$ :

$$\Omega = J^{-1} \Delta J^{-1}$$

with

$$J = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & J_T \end{bmatrix}$$

$$J_t = E \left\{ \frac{\phi_{it}^2}{\Phi_{it}(1 - \Phi_{it})} x_{it} x_{it}' \right\}$$

$$\Delta = E \{ \Psi_{its} \}_{t,s=1,\dots,T}.$$

Each of the  $T^2$  matrices  $\Psi_{its}$  is an  $M \times M$  matrix defined as

$$\Psi_{its} = c_{it} c_{is} x_{it} x_{is}'$$

with

$$c_{it} = \frac{y_{it} - \Phi_{it}}{\Phi_{it}(1 - \Phi_{it})} \phi_{it}$$

All  $\Phi_{it}$  and  $\phi_{it}$  are evaluated in  $x_{it}' \beta_t$  and the expectations are replaced by the sample means, leading to consistent estimates for  $\Omega^{-1}$  and consequently for  $\beta^*$ . Hsiao (1986) or Chamberlain (1984) may be referred to for a proof.

Note that by definition  $\Omega$  is positive definite, which is also a necessary condition for the minimum distance estimator to be calculated. Otherwise, a minimum calculated by setting the derivatives of equation (1) equal to zero may result in finding a saddle point.

By construction  $\Omega/N$  is the covariance matrix of  $\pi^*$ .<sup>2</sup> Therefore it can be derived that  $(F' \Omega^{-1} F)^{-1}/N$  is the covariance matrix of  $\beta^*$ . Here  $F$  is defined as  $\partial f(\beta)/\partial \beta'$ .

We applied this estimation procedure to the dataset which will be discussed below. The first-stage probit estimations were found using the standard software package TSP. The second stage of the estimation was performed with a self-contained computer program, written in Pascal. It is recognised that this minimum distance estimation is likely to be programmed much more conveniently in Gauss, which has a structure designed to handle matrices. Gauss, however, was not available to us at the time of the research.

### 3. Stated Choice Data on Contact Decisions

During 1990 and 1991, at a number of universities in Europe - Amsterdam, Zurich, Vienna, Liverpool and Loughborough (UK) - a survey was held among scientists about communication and travel behaviour. The title of the questionnaire was "University Contact Patterns",<sup>3</sup> and consisted of the following five parts: personal background data; availability and use of communication media; actual contacts (telephone calls and visits); a module with stated choices on contact decisions; and a similar module on media choice. Here we only use the data on the contact decisions; see Button *et al.* (1993) and Ouwersloot (1994) for analysis of other parts of the dataset. In the present analysis we restrict ourselves to the Amsterdam sample, which counted 149 respondents.

The contact decision element consisted of one question, describing a hypothetical situation in which the respondent had to decide whether he would accept an invitation from another scientist or not. Notice that these data are generated using a stated choice (or stated preference) technique, not by revealed preferences. Concerning the use of stated preference data see, for example, the special issue of the *Journal of Transport Economics and Policy* (1988).<sup>4</sup>

An example of the questions raised in the contact decision section of the questionnaire reads as follows:

Imagine that you receive an unsolicited invitation from a *full professor* who you know to be a *leader in your field* to collaborate in the writing of a book. This person is based in *Lyon (France)* and you discover that *his command of English is very*

<sup>2</sup> This is a pragmatic notation. More formally:  $\sqrt{N}[\pi^* - f(\beta)]$  is asymptotically distributed normal, with mean 0 and covariance matrix  $\Omega$ .

<sup>3</sup> This questionnaire was used in the context of the NECTAR working group "Barriers to Communication". NECTAR was an ESF-sponsored research project on Communication and Transport. Besides our own work which produced the Dutch sample, we profited from the efforts of Manfred Fischer and Christian Rammer (for the Austrian data), Rico Maggi (Switzerland), Ken Button (Loughborough), and Peter Brown (Liverpool). The work of the group is reported in Button *et al.* (1993).

<sup>4</sup> The article by Kroes and Sheldon (1988) gives a brief introduction to stated preference techniques, including the potential drawbacks of laboratory behaviour. Wardman (1988) addresses this latter topic in particular, paying specific attention to the external validity of SP techniques.

**Table 1**  
*Varying Characteristics in Contact Decision*

<i>Characteristic</i>	<i>0</i>	<i>1</i>
Place	Lyon (France)	Mexico City
English	Poor	Good
Reputation	Someone you never heard of before	Leader in the field
Position	Junior academic	Full Professor

*poor*. He offers to pay half of the costs of a visit to Lyon — a city that you have already been to in the preceding year — for a face-to-face meeting in order to discuss the proposal. Because of prior commitments on his part the meeting would have to take place during term time and would involve you in substantially rearranging your teaching and administrative duties. Would you go?

The only answers allowed are 'yes' and 'no'. On the basis of this question one can generate more choice sets by varying the characteristics in italic type in a systematic way.<sup>5</sup> The values used for these characteristics are given in Table 1: where there are four of those characteristics which are varied in a binary form. So in all, sixteen ( $2^4$ ) different versions of the question existed, and these were systematically combined in sets of four which were then distributed at random among the respondents. Each set of four questions existed of two pairwise opposite sets, and the two sets again had a maximum Hamming distance. For example, indicating the two values each characteristic can assume by 0 and 1, each version is characterised by a binary vector of length 4. A set then existed, for example, of the versions: { (0,0,0,0), (1,1,1,1), (0,1,0,1), (1,0,1,0) }. Eight such sets were constructed.

Besides these varying aspects there are some characteristics of the choice situation which are kept constant, but are nevertheless highly relevant. These are:

- the respondent is invited, which means it is relatively easy to make the contact;
- the inviter is paying half the cost, thus significantly reducing the monetary cost of establishing the link;
- the respondent is supposed to have visited the city last year. This is intended to reduce the possibly appealing aspect of visiting a city the respondent has never been to before. It also removes some potential barriers of going to a completely unfamiliar place, thus reducing the non-monetary costs;
- a considerable rearrangement of duties is supposed, suggesting a significant non-monetary (time) cost.

The data are analysed in detail in Ouwersloot (1994, chapter 4) and Ouwersloot and Rietveld (1993). In this paper we will concentrate on the effects of the repetition of observations.

<sup>5</sup> The italic type was not used in the original questions.



**Table 2**  
*Evaluation of Chamberlain-Hsiao Estimation*

<i>Parameter</i>	<i>Pooled Data</i>	<i>Subsets</i>	<i>Chamberlain</i>
Constant	-0.79 (0.11)	-0.84 (0.22) -0.71 (0.22) -0.98 (0.28) -0.79 (0.24)	-0.74 (0.12)
English	0.47 (0.11)	0.80 (0.22) 0.30 (0.23) 0.25 (0.23) 0.42 (0.23)	0.44 (0.10)
Place	-0.51 (0.11)	-0.38 (0.24) -0.76 (0.26) -0.41 (0.25) -0.37 (0.24)	-0.37 (0.09)
Reputation	0.69 (0.11)	0.72 (0.24) 0.69 (0.27) 1.04 (0.26) 0.56 (0.24)	0.58 (0.09)

Minimum Distance Statistic for Chamberlain-Hsiao estimation: 20.15

Minimum Distance Statistic of Pooled estimation: 24.09

#### 4. Estimation Results

Table 2 presents the results of the estimations.<sup>6</sup> The first column shows the outcomes when the specific error structure caused by the repetition of observations from identical individuals is neglected. These outcomes are to be compared with those in the last column, from the Chamberlain-Hsiao procedure described in Section 2.

In addition, Table 3 shows the correlation matrix of the errors resulting from the pooled estimation. The table shows that the computed errors are indeed positively correlated,<sup>7</sup> and hence that straightforward application of probit estimation is not justified as this assumes independence of all errors. The interpretation of these positive correlations is that respondents appear to have a certain positive or negative bias in their decision to establish a contact. Those who respond "yes" to one of the questions will also have a higher probability of a positive response to the other questions compared to somebody who responds "no" to the first mentioned question.

Note that the correlation coefficients between cases 1 and 2, and 3 and 4 are considerably less than the others. This most probably stems from the fact that the question-

<sup>6</sup> This table is based on estimations with only three variables and a constant. This specification of the decision is used for the sake of manageability but is certainly not the best specification possible.

<sup>7</sup> Some additional calculations show that this positive correlation is also significant in a statistical sense.

**Table 3**  
*Correlation Matrix of Errors*

	<i>ERR1</i>	<i>ERR2</i>	<i>ERR3</i>	<i>ERR4</i>
ERR1	1.00	0.23	0.48	0.55
ERR2		1.00	0.43	0.47
ERR3			1.00	0.23
ERR4				1.00

naire was constructed in such a way that choice situations 1 and 2, and 3 and 4 were pairwise orthogonal, meaning that, for example, in cases 1 and 2 the variables all had opposite values. Notice that this is in line with the remark in Section 2, that in the present context the correlation coefficient is not necessarily affected by  $(t-s)$  (where  $t$  and  $s$  are indices for case 1-4), but should be modelled in a general way, to allow any form of correlation.

The second column of Table 2 shows the parameters when the four subsets are taken as separate sets.<sup>8</sup> These four vectors  $\beta$  are the basis for the estimation according to the Chamberlain-Hsiao procedure, the results of which are presented in the third column. The bottom row presents the Minimum-Distance test statistic.

The results show that taking account of the correlation structure does have some impact on the estimates. In general, however, this effect is modest: no parameters change signs, or change from being insignificant to significant or vice versa. Moreover, all parameters have a comparable order of magnitude.

In addition to this informal evaluation of the estimation procedure, a formal statistical test can be carried out. According to Chamberlain (1984), the value of the Minimum Distance Function (MDF) at its minimum is a statistic with a  $\chi^2$  distribution. The number of degrees of freedom is equal to the number of constraints that are imposed on the estimates from the subsample estimation. So, in this case, twelve restrictions are imposed (that is,  $\beta_{11}=\beta_{21}=\beta_{31}=\beta_{41}$ , and so on). The value of the MDF equals 20.15 as reported in Table 2. The value of the MDF can also be calculated when four additional constraints are imposed, namely that the resulting estimates are equal to the estimates with pooled data. Thus the number of degrees of freedom is sixteen, equal to the number of initial parameters. Obviously in this case there is nothing to be minimised.

The value of the MDF in the latter case equals 24.09 as shown in Table 2. We can subsequently calculate the difference between the two MDF values. This difference is a test for the null hypothesis that the error structure has no effect on the estimation results.

<sup>8</sup> Each subset consists of one choice situation for each respondent. Recall that the choice situations were randomly distributed (although in fixed, precomposed sets) among the respondents. Hence, each subset shows variation in the exogenous variables.

This null would lead to identical outcomes for Chamberlain-Hsiao and pooled estimation.

The difference has a  $\chi^2$  distribution with four degrees of freedom. Since it is equal to 3.91, the null hypothesis clearly cannot be rejected. Hence, we have to conclude that the specific error structure - in the present case - has no effect on the estimations and that the straightforward use of the probit model causes no problems in this respect.

### 5. Concluding Remarks

In this paper we paid attention to an often neglected effect in the analysis of stated choice data with repeated observations, namely the intrinsically present autocorrelation of the error terms. We discussed the solutions known from the panel data literature and applied an estimation procedure proposed by Chamberlain (1984), also described by Hsiao (1986). The mere application of this procedure proves that it is possible to take account of the correlation structure. The results of the analysis showed, however, that in the present case, which described a contact decision of scientists, the effect is very modest and statistically insignificant.

Clearly, no general conclusions can be drawn from this exercise, concerning the seriousness of the autocorrelation problem in practice. Therefore we advocate the repeated application of methods, like the one described and applied in this paper that take proper account of correlation structure, to gain experience on these phenomena.

### References

- Bates, J. J. (1988): "Econometric Issues in SP Analysis". *Journal of Transport Economics and Policy*, 22, 1, pp.59-70.
- Butler, J. S. and R. Moffitt (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model". *Econometrica*, 50, 3, pp.761-64.
- Button, K., P. Brown, M. Fischer, R. Maggi, H. Ouwersloot, C. Rammer, P. Rietveld and I. Salomon (1993): *Academic Links and Communications*. Aldershot: Avebury.
- Chamberlain, G. (1984): "Panel Data". In Griliches, Z. and M.D. Intriligator (eds): *Hand-book of Econometrics*, vol. II, pp.1247-1318. Amsterdam: North-Holland.
- Hsiao, C. (1986): *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- Kroes, E.P. and R.J. Sheldon (1988): "Stated Preference Methods: An Introduction". *Journal of Transport Economics and Policy*, 22, 1, pp.11-25.
- Maddala, G. S. (1987): "Limited Dependent Variable Models Using Panel Data". *Journal of Human Resources*, 22, 3, pp.303-38.
- Ortúzar, J. de D. and L. G. Willumsen (1990): *Modelling Transport*. Chichester: John Wiley & Sons.
- Ouwersloot, H. (1994): *Information and Communication from an Economic Perspective*. PhD. Thesis, Free University Amsterdam
- Ouwersloot, H. and P. Rietveld (1993): "Formation and Maintenance of Knowledge-Based Networks – The Case of University Contact Patterns". In Bakis, H., R. Abler and E.M. Roche (eds): *Corporate Networks, International Telecommunications and Interdependenc*, pp.161-80. London: Belhaven Press.
- Wardman, M. (1988): "A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour". *Journal of Transport Economics and Policy*, vol. 22, no.1, pp.71-91.

*Date of receipt of final manuscript: November 1994*